

Weekly Report

26/11/2018 – 07/12/2018

I have finished learning how to python programming with NLP by NLTK API. Then I try to figure out a topic for try NLP programming by myself. Anyway, at the same time, my Chinese friend whom I reported to you last time, she asks me about her project to compare the Chinese translated literary works with the original English version (more than 100K words per literature).

That project wants to deal with many indices such as the number of tokens, the number of non-repetitive words, the number of sentences, the number of syllables, the number of letters. These indices can use to find the average word length, the average word length by syllable, the average sentence length, the type/token ratio, the standard type/token ratio, the Flesch reading ease score, the categorizing and tagging words (for content words) and the ratio of content words. I think the indices project can deal with the way I have learned from Youtube. So, I agree to help her with python, and I thought it is a good chance that I could get more understanding the NLP by the real project and learned more about the NLTK API for English and Bosonnlp API for Chinese.

At the end of the work, I have learned more about the back-end process for gathering and preparing data for text visualization like word-cloud. I have learned a lot about how the difference between Chinese and English, for instance, the word tokenizes, and sentence tokenize of Chinese and English, and I just have known there is a readability test designed to indicate how difficult passage in English is to understand and it also can apply for Chinese. Next, she will execute all documents she has then get the data to her professor. Moreover, she will ask her professor for adding me as a co-author, if this point is succeeded, it is a benefit for cooperation.

Plan for the next week, I will back to my track to find a way to extract functional and non-functional requirements and may ask my friend to help how to understand the sentence on social media (use the knowledge of interpreting). Moreover, I will check the Knowledge graph + tag cloud project that you asked me before; perhaps it can adapt to the visualization technique for functional and non-functional requirements gathered from Twitter.